Three Problems with the Empiricist Conception of Concepts

1. Systems properly described as *understanding* something about their environment, for instance that it contains red things, or water, must be able to apply empirical concepts, such as red and water. What must be true of a system for its going into a certain kind of state or producing a certain kind of performance properly to count as the application of such concepts?

There seem to be three ways in which conceptual contents might be attached to states and performances. First, they might be attached *intentionally* to independently specifiable vehicles by the *stipulation* of some interpreter who is antecedently able to deploy the concepts involved. This is how the sign-designs of artificial formal languages come to express the contents they do: what *makes* an expression in such a language mean successor, between, or is red is just that the theorist who uses the language explicitly *takes* it to express those concepts. There is also a familiar line of thought according to which it is in this broadly instrumental fashion that meanings are associated with the expressions of *natural* languages as well: words and sentences mean what they do because linguistic practitioners who are antecedently capable of thought *use* them, by a combination of intention and convention, to convey their concepts and beliefs.

Second, conceptual contents might be attached to states and performances *functionally*, in virtue of their relations to one another and to elements of the environment to which they apply. This approach construes conceptual contents as consisting in the significance states and performances have within the behavioral economy of a system because of their interactions with

each other and with the environment. According to this sort of account, it is the activity of the system itself that establishes the conceptual contentfulness of the states it exhibits, the performances it produces, and the expressions it uses.[1] In this way it is distinguished from the first sort of account, according to which conceptual contents are conferred by the activity of an external interpreter. The functional picture does not require that conceptual contentfulness derives from the antecedently conceptually articulated intentions and beliefs of another. Rather, in some sense the system itself implicitly takes or treats its own states as contentful, and thereby makes them so.

The stipulative and the functional stories are alike in addressing the question of how conceptual contents might be attached to some vehicle that is describable independently of its having the content it has: for instance sign-designs as marks or noises, performances as bodily movements, states as voltage distributions or neurophysiological conditions. It is these vehicles that the interpreter intentionally invests with meaning by stipulation or convention, on the first approach, and that play the functional roles in virtue of which they implicitly acquire such meanings, on the second. A third possibility is that for some items, conceptual contentfulness is *intrinsic,* in that it makes no sense to talk about its being attached to such a vehicle. The conceptual contents of states such as belief, desire, and intention, it might be thought, are *essential* to them. They differ from linguistic expressions such as sentences precisely in that there is nothing that plays the role for them that physically describable sign designs do for

[1] Notice that this abstract characterization allows the possibility that the system whose doings one must consider as conferring genuine conceptual content may have to be a whole linguistic community, rather than a single agent.

sentences, as something independently characterizable to which conceptual content is *attached* (by anyone's activity).[2]

The issues discussed in this chapter arise in connection with pursuit of the second or functionalist explanatory strategy. The stipulative approach is not in principle available as a candidate for explaining the conceptual contentfulness of our thought - or that, if any, of chimps, dolphins, or martians. And if it describes the only way in which conceptual contents can be understood to attach to the states and performances of machines we build, then we will never be able to make them do what we would like them to do, what we can do. For we are semantically autonomous—one need not look to the activity of anyone outside of us in order to make sense of the conceptual contentfulness of our thought and talk.[3] The sort of conceptual contentfulness the stipulative strategy addresses is parasitic on or *derivative* from the more basic sort possessed by the stipulator or interpreter. By contrast, the other two approaches address the sort of *original* conceptual contentfulness our own thought might be taken to have.

The third way of thinking about conceptual contentfulness is put to one side for a different reason: it is not clear how to follow out its consequences for the project of diagnosing or producing concept-use in other creatures, such as chimps, dolphins, and various sorts of hypothetical aliens, nor for the project of artificially constructing genuine concept users.

---

[2] John Searle, in *Intentionality* [Cambridge University Press, 1982] puts forward a view of this sort. Although sentences (in the sense of marks and noises) play an essential role for Davidson, his account of intentional states such as beliefs also has this vehicleless character. Only the whole human subject is the vehicle, and is the identical vehicle of *all* its intentional states.

[3] Again, notice that to say this is not to insist that each of us is individually semantically autonomous in this sense. Insofar as language use is essential to the genesis of genuinely conceptual content, it is only the whole linguistic community—all of us—that exhibits this sort of autonomy. According to such an account, our interpretation of each other might be an essential part of the story; but it will not be interpretation of the sort involved in stipulating, say, that '→' is to be understood as expressing strict implication.

Methodologically, it makes sense to try to specify—in nonsemantic, nonintentional terms—what one has to be *doing* in order to be applying concepts (and hence *understanding)* in an original or nonderivative way. Success would provide the critical criteria of adequacy needed to make progress on the diagnostic, pedagogic, and engineering projects regarding non-human understanders.[4] And attention to the particular ways in which our attempts fall short can in their own way be equally instructive. The rest of this chapter is accordingly addressed to some of the aspects of our activity that have been thought to be of particular significance for the functional conferral of conceptual content on states, performances, and expressions playing suitable roles in that activity.

2. Historically there have been four different approaches to the question of what sort of functional involvements should be understood as conferring conceptual contents:

1) *Empiricism,* of the sort associated with Locke, takes it that the contents of the concepts applied by knowers are a function of their *origin in experience.* This tradition counsels us to look to the *causal antecedents* of concept-use, to the empirical circumstances in which concepts are applied, in order to understand the nature of their contentfulness. The content of a concept is to be read off what its application is typically or correctly a response to. One grasps the concept red, and so has it available to apply to things, only because one has had experiences of red things.

2) *Pragmatism,* of the sort associated with Peirce or James, takes it that the contents of the concepts applied by knowers are a function of *their significance for action.* This tradition counsels us to look to the *causal effects* of concept-use, to the *practical consequences* of

[4] This is the topic of Chapter Eight.

applying concepts, in order to understand the nature of their contentfulness. The content of a concept is to be read off what the typical or correct response is to its application. One applies the concept *food* to something by being disposed to eat it when hungry.

3) *Rationalism,* of the sort associated with Leibniz, takes it that the contents of the concepts applied by knowers are a function of their *role in reasoning.* This tradition counsels us to look to the antecedents and consequents of concept-use, not in the order of *causation,* but in the order *of justification.* The content of a concept is to be read off the good inferences in which it serves as premise or conclusion. Grasping a concept is mastering its inferential role—knowing for instance that being a trilateral rectilinear plane figure is a sufficient condition for the applicability of the concept triangle, and that having angles adding up to a straight line is a necessary condition of its applicability.

4) Finally, the *synthetic or combined* view, which may be associated with Kant, takes it that the contents of concepts involve all of these dimensions. According to this approach, the challenging theoretical question is not which of these to focus on—the causal or the inferential, circumstances or consequences of application—but rather how these various aspects of concept use should be understood to be related to one another. What is the nature of the relation between causal and inferential grounds of application of an empirical concept, or between its causal and inferential consequences? What underwrites the transition from its circumstances of application (of either sort) to its consequences of application (whether practical or theoretical)?[5]

[5] I am here suppressing some issues that are important in other contexts. For one can separate the question of whether one should focus on perceptual experience, inference, or action from whether one construes those aspects of concept-use in causal or normative terms.

The synthetic view has the familiar agreeable virtues of a position that resolves a dispute by allowing the participants to have their cake and eat it too. There are all sorts of concepts, and no doubt all sorts of functional relations are involved in their articulation—both inferential relations to other concepts and noninferential causal relations with elements of the environment, through both perception and action. Important insights concerning the contributions of these different sorts of functional relations are in danger of being obscured, however, if we rush to embrace an eclectic approach without looking at what each of the more partial views it embraces can and cannot illuminate. In particular, I believe that the rationalists are right to the extent that to talk about a state, performance, or expression as *conceptually* contentful is to talk about it as playing a distinctively *inferential* role. It is its inferential articulation that is distinctive of the conceptual as such. It is only within the framework of specifically inferential involvements that *causal* interactions with the environment—whether upstream, as in perception, or downstream, as in action—can contribute to specifically conceptual contents. Our ordinary concepts do involve both empirical and practical dimensions, but no amount of empirical or practical functional articulation can take the place of the inferential functional relations characteristic of concepts.

This point is not always appreciated, and in order to reinforce it, I want in the rest of this chapter to consider three systematic sorts of explanatory failings to which attempts to specify conceptual contents purely in terms of causal inputs and outputs are subject. I will start with empiricist attempts to elaborate conceptual contents on the basis of a system's reliable dispositions to respond differentially to various sorts of stimuli, since I think this remains the

most tempting sort of reductionist ploy. But I will indicate in each case why supplementing the empirical inputs with practical outputs does not obviate the difficulties being considered.

3. We may start with a very simple question. What more is there to grasping the empirical concept red than being able to respond differentially to red things? The empiricist intuition is that what makes something a representation of red is its status as a response that is reliably differentially elicited by red stimuli. A suitably frequency-sensitive photocell can do that, or a well-trained pigeon. It seems, however, that a distinction ought to be made between the capacity merely to respond differentially to red stimuli and the capacity to respond differentially to them by applying the *concept* red. After all, a chunk of iron responds differentially to its environment by rusting in, and only in, the presence of water and accelerating in, and only in, the presence of forces. Indeed, every actual object produces behavior of different kinds in response to stimuli of different kinds.

Although such differential responsiveness is no doubt a *necessary* condition for mastery of empirical concepts, it is not at all plausible to take it as a *sufficient* condition. Stones and photocells certainly do not apply concepts when they respond differentially to impressed forces or energetic photons. And although much of the psychological literature whose topic is labeled "concept formation" in fact addresses little else than this sort of reliable differential recognitive capacity, it requires quite a stretch to say that pigeons acquire the *concept* red simply by learning to peck at a round button when presented with a red stimulus.[6] The empiricist is not in a position

[6] And it is only in the context of the stipulative paradigm for attaching conceptual content to performances that it makes any difference whether the round button bears the sign design 'red', or the photocell is wired up to produce a tokening of that type on a VDT.

to say what discriminates specifically *conceptual* responses or representations from simpler *non*conceptual discriminations.

At this point it is tempting to invoke *complexity.* There is a sliding scale, it will be claimed, with stones and photocells at the bottom, simply trained pigeons somewhat farther up, more complexity trained chimps further up, barely linguistically capable human children above them, and linguistically sophisticated human adult reporters of the colors of stoplights and stop signs at the top. Our resistance to taking the discriminative capacities exhibited at the lower end of the scale as involving the application of concepts is just a matter of their insufficient complexity.

In some sense this suggestion must be correct. But it is nonetheless almost completely unhelpful with respect to the point at issue. For there is complexity and complexity. Pigeons are not *biologically* substantially less complex than we are. Everything turns on just what *sort* of complexity of differential response is sufficient for genuine concept use. It is pretty clear that neither increasing the fineness of discrimination nor the sheer numbers of stimulus-response connections can by themselves suffice to turn the individual responses into applications of concepts. Intuitively what is missing in these cases is *understanding.* The responses of a genuine concept-using reporter of red *mean* something to that reporter in a way that the responses of photocells and pigeons do not. (The responses of the photocell and the pigeon can mean something to *us,* who can attach conceptual significance to them—but this is relevant to the stipulative strategy, not the functionalist one). What sort of 'complexity' of differential

responsive disposition corresponds to *understanding* (hence to sapience)? The empiricist intuition about contentfulness does not supply the resources for an answer.[7]

Nor is it clear what sort of additional help is offered by adding the considerations advanced by the pragmatist concerning practical activity - looking at outputs rather than just inputs. There is a sense in which an animal treats something as food by eating it. For what it produces is not just any differentially elicited response, but one that is appropriate to it as food. And this is a proper functional matter of the role the stimulus plays in the activities of the animal itself, not simply a matter of our stipulation or interpretation. But doesn't the stone also respond to an impressed force as an impressed force by producing the response that is appropriate to it as such, namely accelerating? And doesn't the pigeon's training simply create an environment in which the appropriate response for the pigeon to something *as* red is to peck at the round button? Again, nothing seems to mark off the distinctively conceptual.

One could at this point throw up one's hands, jettison the functionalist project, and concede that what is missing can only be supplied by invoking *intrinsically* conceptually contentful states of understanding—something that we concept-mongering sapients have, but that mere sentients such as pigeons do not. But there is an alternative. The rationalist's idea is that what is distinctive of the conceptual is its specifically inferential articulation. For a response reliably differentially elicited by red things to qualify as the application of a concept, it must have the significance of a move in the game of giving and asking for reasons. It must be available as a premise for drawing further conclusions, and be liable to challenge by inferences from premises with incompatible consequences. Grasping the concept red requires practical

[7] This issue is discussed in more detail in Chapter Eight, where a specific, four-leveled hierarchy of semantic complexity of concepts is described.

mastery of its *inferential role,* as well as its *non*inferential role as an appropriate response to certain sorts of stimuli. This means being able to discriminate what follows from it, what it follows from, and what is incompatible with it—endorsing in practice the moves from red to colored**,** and from scarlet to red**,** and treating red as ruling out green. (Indeed, the considerations advanced by the pragmatist can be seen to be subsumed by those advanced by the rationalist, for the inferential role of a claim or concept includes its premissory service in *practical* inferences, those whose conclusions are intentional actions.)

When we say that the reliable differential responses elicited by red things don't *mean* anything to the photocell or the pigeon—that they don't *understand* anything by those responses, and hence are not grasping or applying concepts—we are pointing to the fact that they don't treat those responses as *inferentially* significant, as premises providing potential evidence for further conclusions and as potentially confirmed by or clashing with the conclusions that follow from other evidence. When we say that in spite of the conceptual incapacity of photocells and pigeons it is nonetheless possible for *us* to interpret their responses conceptually—*to attach* conceptual content to their otherwise nonconceptual differential responses—we are saying that *we* can use the fact that they responded in a particular way as a premise for inference, confront it with incompatible evidence, and so on. This is in effect using them as measuring instruments, exploiting their reliability by endorsing inferences from claims about their responses ("The pigeon pecked the round button,") to claims about the stimuli that elicited those responses ("So the ball it was looking at was probably red,"). Thus the rationalist insight—that conceptual content is attached to states, performances, and expressions by the role they play in *reasoning*—provides an answer to the demarcational question that empiricism and pragmatism cannot

answer. It potentially offers the missing ingredient needed to turn into *sufficient* conditions the *necessary* conditions for grasp of empirical concepts that they provide.[8]

4. A second dimension along which the empiricist construal of concept-use can be seen to be in need of supplementation by considerations of the sort forwarded by the rationalists concerns the notion of *reliability.* According to the empiricist, being able to apply a concept such as red consists in being able reliably to respond differentially to red things, for instance by saying "That's red'. Systems that cannot do this reliably don't count as grasping the concept. The reliability of discriminative capacities, the criterion of concept-use, is understood as an objective matter of fact. In particular, it consists in the objective probability of a correct response. With respect to the dispositions of the system at a given time, this probability can be identified with the relative frequency of correct responses, the proportion of circumstances in which the system is disposed to respond correctly. in this way the sort of understanding involved in taking something as something by applying a concept to it is *naturalized,* by identifying it with an aspect of the purely cause order.

The difficulty with this aspect of the empiricist approach is, briefly, this. The responsive dispositions of a system determine the objective probability of, say, producing the response appropriate to the presence of a red barn when (or only when) stimulated by a red barn, only relative to a reference class of possible situations. No matter what its dispositions, any system

[8] Notice that to say this much is not yet to say what is required for specifically inferential articulation. It might be understood simply as more conditioned stimulus-response connections among states and performances - a matter of being disposed to make one move in response to others. At the least such a story must include an account of what it is for such a 'move' to have the status of a claim or judgment, taking a stand, undertaking a commitment, that one can come to be entitled to in various ways and that can commit and entitle one to other such stands. What can in this way both serve as a reason for other stands, and is itself liable to demands for reasons, will be recognizable in virtue of that role as having a specifically propositional content, the fundamental sort of conceptual content.

will count as reliable relative to some reference classes, and as unreliable relative to others. But the causal order itself does not privilege any of these reference classes over any other; there is no objective way to pick out one or some of them as correct. Thus there is no objective fact of the matter as to whether a given system is a reliable discriminator of red things. Accordingly, whether or not a system counts for the empiricist as grasping and applying concepts such as red barn is not—as the empiricist wants it to be—settled by its responsive dispositions.

To see this, consider a version of an example put forward in another context by Goldman.[9] Consider Barney, whom we would intuitively take to be a reliable reporter of red barns. Barney is looking at a red barn in good light, and he reports "There's a red barn." What Barney doesn't know, however, is that he is in Barn-Façade County. The local hobby in that county is building incredibly lifelike *trompe l'oeil* façades of red barns. 99% of what appear to be red barns in this county are actually such façades. Barney is not capable of discriminating real barns from these façades. It is just dumb luck that he happens to be looking at one of the few real red barns in the county. Thus as he wanders around the county, Barney will be wrong 99 times out of a hundred in his barn-reports. However, these façades are the *only* thing that Barney can't tell from real red barns, and the county is small enough that those amount to only 1% of the apparent barns in the whole state—the rest of whose inhabitants spend their time more productively than the façade-fanciers. Within the state as a whole, as a consequence, Barney's reports are right 99 times out of a hundred.

[9] Alvin Goldman, "Discrimination and Perceptual Knowledge", *Journal of Philosophy,* vol. 73, no. 20 (1976).

According to the empiricist criterion, Barney has the concept of red barns just in case he has a reliable disposition to respond differentially to them with a red-barn report. Does he or doesn't he? If we draw the boundaries around the relevant reference class narrowly enough, then he is reliable. For in the case described he is looking at an actual barn, and so with respect to *that* situation he is 100% reliable. Broadening the boundaries to include the whole county reveals him as unreliable, however. But if the boundaries of the reference class are broadened still further, to the whole state, his reliability is, so to speak, reinstated. Clearly cases could be described in which such oscillations are repeated indefinitely. Equally clearly, no simple rules about always using the narrowest or the widest available reference class will resolve the difficulty.

It would be bad enough if we had to say that Barney gains and loses his mastery of the concept red barn as he moves in and out of Barn-Façade County. But Barney isn't even moving. Contradictory verdicts on his grasp of that concept right now result from assessing his reliability with respect to different reference classes. Yet his place in the causal order, which includes all his responsive dispositions, doesn't settle the question of which is the correct reference class, and so doesn't settle whether or not he is reliable. Reliability is not in objective matter of fact; for it depends on the choice of a reference class, and that choice is not determined by objective matters of fact.

Again the rationalist emphasis on the essential role played by mastery *of inferential* roles in the grasp of concepts supplies what is missing from the empiricist account. First, wherever Barney is, and no matter how a reference class is chosen in accordance with which to assess the

reliability of his noninferential responsive dispositions, he can count as grasping the concept red barn in virtue of his knowing what follows from applying that concept—e.g. that it is a structure the color of stop signs and sunsets, in which farm animals can live—and what other claims would be evidence for or against something qualifying as a red barn. Indeed, it is precisely his mastery of these aspects of the inferential articulation of the concept that could lead him to discover that his noninferential responsive dispositions were leading him astray in Barn-Façade County.

This possibility of Barney's using his mastery of the concept to discover his local *un*reliability as a noninferential reporter points to the second way in which the rationalist perspective sheds tight on the intuition that underlies the empiricist appeal to reliability. For to take Barney to be a reliable reporter just is to endorse an *inference*—the inference, namely, from Barney's responding to something *as* a red barn to its *being* a red barn. To take a certain instrument to be a reliable indicator of the current in a wire is to endorse the inference from its meter-needle indicating a current of 5 amps to there being a 5 amp current running through the wire. Indeed, if Barney discovers his plight, he will himself cease to endorse the inference from 'I am in Barn-Façade County and am non- inferentially disposed to report the presence of a red barn,' to "There is a red barn in front of me." (As Sellars[10] has taught us, at this point Barney can express his responsive disposition while withholding his endorsement of the claim he is tempted

[10] *Empiricism and the Philosophy of Mind* by Wilfrid Sellars, Robert B. Brandom (ed.) [Harvard University Press, 1997].

to make by his unreflective dispositions by saying “It *looks* as though there is a red barn in front of me.")

The fact that assessments of reliability are relative to choice of reference class finds straightforward accommodation in this rationalist picture. It simply reflects the fact that endorsement of the reliability inference—from a system's responding to something as a **K** to its being a **K**—is sensitive to auxiliary hypotheses. The inference from “Barney is looking at a red barn and is noninferentially disposed to report the presence of a red barn,” to “There is a red barn in front of Barney,” is a good one. The inference from “Barney is in Barn-Façade County and is noninferentially disposed to report the presence of a red barn,” to "There is a red barn in front of Barney," is not a good one. The inference from “Barney is somewhere in the state and is noninferentiallv disposed to report the presence of a red barn,” to “There is a red barn in front of Barney,” is a good one. The fact that in the situation originally described the premises of each of these three inferences are true is why his reliability can be variously assessed, depending on how his situation is specified. The difference between the two approaches is that the rationalist can, as the empiricist cannot, accommodate these different assessments of reliability without being obliged to make mysterious and even contradictory claims about Barney's mastery of the empirical concept red barn.

5. Another area in which the empiricist construal of concept-use can be seen to be in need of supplementation by considerations of the sort forwarded by the rationalists has to do with the identification of what is reported, what a concept is being responsively applied *to*. The stimulus-response model of empirical conceptual content faces a fundamental problem concerning the

discrimination of the *stimulus.* In his discussion, Davidson introduces the familiar point this way:

> The location of a stimulus is, of course, notoriously ambiguous. We can place it almost anywhere in the causal chain that leads from far outside to various parts of the central nervous system. Quine offers us a choice between two of the possible locations: at the sensory receptors, or at the object and events our observations are traditionally about ... it makes a vast difference whether meaning or evidence are tied to the proximal or the distal stimulus. Mindful of a certain tradition, let us call the two resulting theories of meaning and evidence the *proximal* and the *distal* theory.[11]

In another place, this difficulty is explained further:

> ... why say the stimulus is the ringing of the bell? Why not the motion of the air close to the cars of the dog—or even the stimulation of its nerve endings? Certainly if the air were made to vibrate in just the way the bell makes it vibrate it would make no difference to the behavior of the dog. And if the right nerve endings were activated in the right way, there still would be no difference. And in fact if we must choose, it seems that the proximal cause of the behavior has the best claim to be called the stimulus, since the more distant an event is causally the more chance there is that the causal chain will be broken.[12]

Typically, there is a whole causal chain of covarying events culminating in a response. In the standard case, the occurrence of one is accompanied by the occurrence of all the rest. Under

---

[11] "Meaning, Truth, and Evidence" (hereafter *MTE*, in *Perspectives on Quine* R. Barrett and R. Gibson, eds. (Oxford: Basil Blackwell, 1990), pp. 68-79.

[12] "The Conditions of Thought", *Le Cahier du College International de Philosophie,* Editions Osiris, Paris, 1989, pp. 165 - 171, reprinted in *The Mind of Donald Davidson, J.* Brandl and W. Gombocz eds. Edition Rodopi *(Grazer Philosophische Studien Band 36)* 1989 pp. 193 - 200.

these circumstances, the response being keyed to one of the event kinds is its being keyed to all the rest. How is one element of the chain to be singled out as the stimulus? What is the nature and source of the privilege that distinguishes one element from another?

One strategy for assigning such privilege, and therefore picking out as the stimulus one element from the whole chain of covarying event types that culminates in a response of the specified type is, as Davidson suggests, to look to *proximity* to the eventual response. Such a proximal theory of stimuli will always yield the result that the stimuli being responded to are at the sensory surfaces or within the nervous system of the responding organism. The justification for seizing on causal proximity of stimulating event to the response as what matters is, as Davidson suggests, maximizing the relative *reliability* of the connection between the occurrence of events of the distinguished stimulus type and the occurrence of events of the distinguished response type. The proximal element of the chain is the one that most reliably brings about the response, since prior occurrences in the chain elicit the response only in the cases where they succeed in bringing about the an event of the proximal type, while events of that type can elicit the response regardless of whether they have themselves been brought about in the standard way.

In the context of the project of using reliable differential responsive dispositions as a model to understand the application of the most basic empirical concepts, the adoption of such a policy for the discrimination of stimuli is the continuation of Cartesianism by other means. [13] The clearest evidence of this is the room that proximal theories leave for global skepticism. Discriminating stimuli by proximity inevitably raises the possibility of skepticism about anything

[13] *MTE: "*Proximal theories, no matter how decked out, are Cartesian in spirit and consequence."

outside the responding organism, since the stimuli being responded to could, on the proximal reading, be just as they are regardless of variations in, or even the total absence of, the world of environing objects. The attempt to maximize reliability in responsive classification of stimuli (the primitive empiricist version of the application of concepts of observables) by contracting to the most proximal stimuli taken to be responded to ends by making the correctness of such classification bear only an ultimately accidental relation to anything happening outside the responding organism.

Thus proximal approaches to distinguishing stimuli are disastrous for the project of understanding the contents of fundamental empirical concepts from the way in which their application in observation depends on the exercise of reliable differential responsive dispositions. For the stimuli that are classified by responses do not extend out to the world, but remain in the organism, or at best at its functional surface. Thus what is classified by the proto-concepts that repeatable responses are going proxy for is not bells and tables and rabbits, but only states of the responding organism. Nothing that looks like one of our ordinary empirical concepts, applying to ordinary observable objects, is within reach of such an approach. A distal strategy is required in order to get the proto-concepts represented by reliably differentially elicited noninferential response types to be intelligible as classifying, and so applying to, ordinary observable objects and properties. Understanding them this way involves respecting the language-learning situation in which these reliable differential responsive dispositions are established.

The most popular approach to identifying distal stimuli as what is classified by the exercise of reliable differential responsive dispositions is to appeal to *triangulation.* This is a

strategy for picking out or privileging one bit of the causal chain of covarying event types that reliably culminates in a response of a distinguished type, by looking at the *intersection* of *two* such chains. The insight it develops is that the best way to pick a single *point* (the stimulus) out of a *line* (the causal chain of covarying event-types that reliably elicit a response of the relevant type) is to *intersect* it with another line--another causal chain corresponding to another reliable differential responsive disposition.

One writer who employs such a triangulation strategy to address the problem of picking out distal stimuli as what a response is about is Fred Dretske, in *Knowledge and the Flow of Information.*[14] He looks to the upstream intersection of two distinct "flows of information", or causal chains of reliably covarying event-types, that reliably culminate in responses of the same type, in order to pick out the distal stimulus. A simple example of the sort of system he has in mind would be a thermostat that keeps the temperature of a room within a certain range by turning a furnace on and off. if the thermostat has only one way of measuring temperature, for instance by the bending of a bimetallic strip until it touches either the left electrical contact (too cold) or the right one (to warm). There is no way, Dretske acknowledges, to say that what the system is responding to is the temperature of the room, rather than the temperature of the bimetallic strip, or the curvature of the bimetallic strip, or the closing of the circuit between the bimetallic strip and one or the other of the contacts. Notice that a pragmatist appeal to *consequences* of the response in question is of no help here; for turning the furnace on affects not only the temperature of the room, but also that of the bimetallic strip, its curvature, and so its relation to the electrical contacts.

[14] MIT Press (Bradford) 1981.

One can be entitled to such a description, however, if the thermostat is slightly more complicated, and has another causal route to the same response (turning the furnace on or off). If the thermostat has a second sensor, for instance a column of mercury supporting a float with an electrical contact that completes one circuit to rum the furnace on whenever the float is below one point (too cold) and turns it off whenever the float is above another point (too warm), then the system has two ways of responding to the change in temperature in the room.[15] Although for this second route by itself, just as for the first by itself, there is no feature of the system that entitles one to say it is responding to changes in the temperature of the room rather than to the temperature of the mercury, or the length of the mercury column, or the closing of the switches, when the two routes are considered together, they intersect in just two places—upstream at the change of temperature in a room which is included in the "flow' or causal chain corresponding to each route, and downstream in the response of turning the furnace on or off. (Since the two routes intersect in two places, it is necessary to think of arcs rather than lines as intersecting, and the term 'triangulation' may be less happy.) But Dretske shows how the general strategy of looking to the *intersection* of two reliable differential responsive dispositions might be funded from the resources of the responding system itself.

One might worry that Dretske has not in fact succeeded in responding to the general worry about how to justify describing the system as responding to a distal stimulus rather than a

[15] For the purposes of the illustration, one need not worry about how the system deals with conflicts arising when the two subsystems disagree. The range of response of one could be set wider than the other (permitting a lower temperature before a 'too cold' message is processed and a higher one before a 'too warm' message is processed), so that the wider one counts as an overriding failsafe or backup, or the control could pass randomly from one to the other in case of disagreement. The exact arrangement would matter a lot for how well the thermostat maintained a steady state temperature within its specified range, but does not seem to affect the triangulation strategy in play.

proximal one. For there is an objection available to his strategy that seems to reinstate the original worry. Why, it might be asked, ought we not to conclude that even in the two sub-system case, what is responded to is a proximal stimulus, but a *disjunctive* one? The system turns on the furnace just in case *either* the temperature of the bimetallic strip is low enough *or* that of the mercury column is low enough, or alternatively, in case the curvature of the bimetallic strip is far enough to the left *or* the mercury column is short enough. (Again, pragmatic appeal to consequences of entering this state won't solve the problem.)

This worry is connected to the complaint voiced earlier, to the effect that mere differential responsiveness is not sufficient for identifying the responses in question as applications of *concepts.* The rationalist supplementation suggested there—that what is distinctive of the conceptual is *the inferential* role played by the responses that stimuli differentially elicit—is also what is required to exploit the triangulation strategy in connection with genuine concepts in a way that responds to the worry about disjunctive proximal stimuli.

Consider an individual who reliably responds, as one wants to say, to the visible presence of rabbits, by saying 'gavagai'. Suppose further that he is reliably differentially responding not just to rabbits, but to the presence of the distinctive (according to him) rabbit flies that are for him decisive evidence of the presence of rabbits, or that the visual cue he is using, as determined by a physiologist of perception, is a glimpse of the fluff around the tail of the rabbit. What is it about the situation in virtue of which he can be said nonetheless to be reporting not the presence of the rabbit flies, or of the fluffy tail, but the presence of a rabbit? The rationalist response is that the difference is not to be found in the reliable differential responsive dispositions, not in the

causal chain of covarying events that reliably culminates in the response 'gavagai', to which not only the rabbit but the flies or the fluffy tail belong, but rather in the inferential role of the response 'gavagai'. For instance, does the commitment undertaken by that response include a commitment to the claim that what is reported can fly? Or is the claim expressed by 'gavagai' incompatible with the further characterization of the item reported as flying? If it is incompatible, then it is not the flies that are being reported. What determines which element of the causal chain of covarying events that reliably elicit the report is being reported is the *inferential* role of the report, what it *entails,* what is *evidence* for it, what it is *incompatible* with.

Assuming that the observable predicate corresponding to 'flying' has already picked out the things that fly, noticing that the report 'gavagai' could mean rabbit flies in case its applicability entails the applicability of 'flying' and could not mean rabbit flies in case its applicability is incompatible with the applicability of 'flying' is just what is wanted to pick out the distal stimulus the concept expressed by 'gavagai' is being applied to or is classifying. But the appeal to inference and incompatibility may seem just to put off the issue. How does 'flying' get to apply properly to flying things, and not to whatever cues we in fact use in discriminating flying things, in short to one element of the causal chain of covarying event types that reliably culminate in its application? The answer must be that what the appeal to inferential role does is establish a sort of *triangulation,* or intersection of flows of information or reliable differential responsive dispositions. If 'gavagai' is used so as to entail 'flying', then whatever is properly responded to by the one expression must be properly responded to by the other, so what is classified as gavagai must also be classified as flying, so 'gavagai' must apply to rabbit flies, and not to the rabbits that are their invariable (we are supposing) concomitants. In short, the

rationalist appeal to inferential role, in addition to reliable differential responsive dispositions, involves triangulation of the sort that Dretske invokes, where two (or more) different reliable responsive dispositions of the system are invoked, so that their intersection can pick out a unique element of the causal chain of covarying events as the stimulus being classified by a response of a certain type. Because 'flying' will *not* be taken to apply to lots of things that merely hop, we can be sure that it does not mean flying or hopping, and so that 'gavagai' does not mean something disjunctive like rabbit or rabbit-fly.

Thus the rationalist view can incorporate Dretske-style triangulation, by using the inferential articulation of the response to connect one observable, with its chain of 'upstream' antecedents, with another chain, with its different but intersecting set of antecedents. It is in this way that concepts such as 'red' and 'colored' work, where each has noninferential circumstances of appropriate application, as well as an inferential link to the other. In languages without purely theoretical claims (that is, claims whose only circumstances and consequences of application are inferential), all concepts in this way answer to two masters: their own noninferential circumstances of appropriate application and their inferential links to the noninferential circumstances of application of other concepts. Dretske could keep his basic insight, while avoiding the embarrassment of not being able to distinguish fancy thermostats from genuine concept deploying believers, if he insisted on inferential articulation along with the intersecting chain move.

Thus to make the triangulation approach to distinguishing distal stimuli work, one needs to look further 'downstream' from the response, as well as 'upstream'—just as orthodox

functionalism would lead one to expect. What picks out one kind of thing as what is being *reported* out of all those that are being differentially *responded* to is a matter of the inferential commitments that response is involved in. These inferential consequences of going into a state make it clear that what is being classified is something outside the system. They are what determine that a physicist is reporting the presence of a mu meson in a bubble chamber, and not simply a large hook-shaped pattern. For the consequences of classifying something as a microscopic mu meson are quite different from those of classifying something as a macroscopic hook-shaped trace. It is the lack of such consequences that makes Dretske's dual thermometer liable to a disjunctive proximal interpretation.

6. At the outset, I distinguished four different ways in which a functionalist approach to the conferral of conceptual content might be pursued: the empiricist strategy, which looks exclusively to the causal antecedents that reliably elicit a (therefore) contentful state; the pragmatist strategy, which looks exclusively to the causal consequences reliably elicited by a (therefore) contentful state; the rationalist strategy, which looks to the role in reasoning of a (therefore) contentful state; and the combined strategy, which appeals to all these sorts of functional involvements, both causal and inferential. I don't take it to be surprising that the combined strategy is the most promising of these, and demonstrating that point has not been my primary goal here. My aim has been rather to show some of the ways in which the rationalist criterion of demarcation of the *conceptual* in terms of specifically *inferential* articulation remedies specific shortcomings of empiricist and pragmatist accounts of conceptual content, which appeal only to causal inputs (in perception) and outputs (in action). To that end, I have considered three difficulties that confront such noninferential conceptions of conceptual content:

inability to distinguish the sort of complexity distinctive of the *conceptual* as such, failure to make sense of the crucial notion of responsive *reliability,* and liability to *proximal* (including *disjunctive*) interpretations of what empirical concepts are applied to. In each case, I have claimed, the rationalist insight supplies what is missing from the empiricist approach.